Int. J. Anal. Appl. (2025), 23:112

International Journal of Analysis and Applications

University Rankings Prediction Using Hybrid Feature Selection Based on Machine Learning Methods

Kittipol Wisaeng[1,*], Benchalak Muangmeesri[2]

[1]*Technology and Business Information System Unit, Mahasarakham Business School, Mahasarakham University, Mahasarakham 44150, Thailand*

[2]*Engineering Management, Suan Sunandha Rajabhat University, 1 U-Thong nok Road, Dusit, Bangkok 10300, Thailand*

**Corresponding:* Kittipol.w@acc.msu.ac.th

Abstract. This study presents a novel approach to predicting university rankings using hybrid feature selection and machine learning techniques. It identifies critical performance factors that affect ranking accuracy using the Times Higher Education (THE) dataset, which includes data from 1,904 universities. A Max-Min normalization method and an artificial neural network were applied to preprocess the data. Then, a hybrid feature selection method, combining statistical and machine learning techniques, was used to determine the optimal feature subsets. Several prediction models, including linear regression, random forest, and multilayer perceptron, were evaluated and compared based on various metrics: accuracy, precision, mean absolute error (MAE), root mean square error (RMSE), and R^2. The results indicate that hybrid feature selection using machine learning significantly enhances predictive accuracy. The hybrid model consistently outperformed all other models across various metrics, achieving the highest accuracy (0.971), precision (0.985), recall (0.971), and F1-score (0.972). These results demonstrate that the hybrid model effectively balances true positive and false positive predictions while minimizing errors. Furthermore, the error metrics for the hybrid model were the lowest among all models, with an MAE of 0.034 and an RMSE of 0.028. This reinforces its superiority in delivering highly reliable predictions. This study demonstrates the effectiveness of hybrid feature selection in refining ranking systems and offers a robust framework that can be applied to various datasets and ranking environments. The findings provide valuable insights for improving ranking predictions and shaping strategies in higher education.

Received Feb. 6, 2025

2020 Mathematics Subject Classification. 68T05.

Key words and phrases. machine learning; hybrid feature selection; predictive accuracy.

https://doi.org/10.28924/2291-8639-23-2025-112

ISSN: 2291-8639

1. Introduction

University rankings have become vital benchmarks for evaluating the quality of higher education institutions. These rankings significantly influence decisions regarding student enrollment, funding allocation, and strategic partnerships within academic and industry sectors. Ranking systems analyze various parameters to assess universities effectively, including teaching quality, research output, citation impact, international outlook, income, and student enrollment. These factors contribute to a comprehensive understanding of a university's overall performance and reputation in the educational landscape. Prominent ranking organizations, such as Times Higher Education (THE) and QS World University Rankings, use different methodologies to highlight various aspects of academic excellence and institutional effectiveness. For example, THE recently introduced the WUR 3.0 framework, representing a significant evolution in calculating university rankings. Furthermore, these rankings utilize extensive datasets to inform their evaluations. The latest analysis from THE scrutinized over 134 million citations from 16.5 million research publications, providing a solid foundation for their assessments. This extensive data collection allows ranking systems to offer stakeholders— including students, faculty, policymakers, and potential collaborators—actionable insights to guide their decisions regarding educational and research priorities. Through this approach, university rankings serve as powerful tools that illuminate the evolving landscape of global higher education and encourage institutions toward continuous improvement and excellence ([1]). Despite their importance in evaluating institutional quality and performance, traditional ranking methodologies face significant limitations in accurately predicting future university performance. These conventional methods often rely on static metrics and historical averages, which may fail to capture the dynamic nature of educational institutions. Consequently, there is a pressing need to adopt more advanced analytical approaches, particularly those based on machine learning. Machine learning techniques analyze historical data to identify patterns and predict future trends.

We aim to overcome the shortcomings of traditional university ranking methods by utilizing advanced algorithms. These advanced statistical techniques are designed to improve the precision and applicability of ranking systems in higher education. A key component of this process is featuring selection, which is essential for developing effective ranking prediction models. One notable study ([2]) introduced an innovative method that integrated the Preference Ranking Organization Method for Enrichment Evaluation (PROMETHEE) with a rough set theory. This approach is designed to quantitatively rank features based on their relevance and contributions to the overall ranking process. Through experiments conducted on diverse datasets, this method demonstrated considerable effectiveness in resolving multi-source ordered decision-making challenges, which are standard in ranking evaluations. Another research effort ([3]) investigated the relationship between university rankings and theoretical frameworks utilized in doctoral capstone projects. This study applied machine learning techniques to compute cosine similarity scores between various capstone projects and established learning theories. The findings of this investigation revealed notable discrepancies in the utilization of theoretical frameworks across different ranking groups. However, the study was primarily limited by its exclusive reliance on datasets from U.S. universities, which may not be representative of global trends. In a focused analysis of Canadian universities ([4]), researchers identified critical ranking factors such as the student-faculty ratio and the total number of academic citations. They used advanced feature engineering techniques, such as Pearson correlation and Chi-Square analysis, to identify the key factors affecting rankings. This study demonstrated how machine learning could enhance ranking predictions and guide university strategies. Additionally, graph neural networks (GNNs) have been developed as a powerful tool for preference learning within the context of university rankings ([5]). By constructing intricate preference relation graphs and employing edge classification techniques, GNNs offer a robust framework capable of enhancing preference ranking accuracy and efficiency. Furthermore, a meta-heuristic approach known as Ranking Improved Teaching-Learning-Based Optimization (RITLBO) has also been explored. This method successfully identified key parameters for performance modeling in photovoltaic systems, illustrating its versatility and potential application in various ranking-related tasks beyond education ([6]). Through these advanced methodologies, the study aims to create more effective and reliable university ranking systems that accurately reflect university performance and support strategic improvements. A comprehensive study ([7]) introduced a sophisticated framework known as CCRank, explicitly designed for enhancing learning-to-rank processes by applying evolutionary algorithms (EAs). This innovative framework leverages the concept of cooperative coevolution (CC), which allows for decomposing complex ranking challenges into more manageable sub-problems. By facilitating parallel optimization strategies, CCRank significantly enhances computational efficiency and effectiveness. The framework has been implemented with three distinct EA-based algorithms, showcasing its ability to achieve high levels of accuracy and scalability. This positions CCRank as a powerful tool for tackling intricate ranking tasks across various domains. In a separate context, the Indian higher education system, despite its expansive size and potential, grapples with numerous systemic challenges deeply rooted in its historical colonial legacy and the realities of its status as a developing country. Although the government has committed to establishing world-class universities, systemic reforms have not progressed to an adequate level. This fixation on achieving high standings in global university rankings has led to benchmarking initiatives that frequently neglect the socioeconomic conditions prevalent in the country ([8]). Comparative studies evaluating THE, QS world university rankings, and the Academic Ranking of World Universities (ARWU) have shed light on the varying methodologies employed in these rankings and the implications these differences hold for various stakeholders in the academic community ([9]). Furthermore, a study ([10]) examined the dynamic interplay between key factors such as academic pedigree (educational background), cognitive ability (often measured as IQ), and cultural intelligence (CQ) within the context of global virtual teams. The research findings indicated that the most effective team configurations combined high IQ and CQ levels with moderate academic pedigree diversity, ultimately enhancing team performance. Additionally, for nearly two decades, scholarly literature has consistently advocated adopting government funding models aligned with the metrics used in global rankings. However, a thorough analysis of performance-based funding across countries such as Australia, Denmark, Finland, and Sweden revealed that there is only a limited alignment between the volume of funding allocated and the criteria set forth by these ranking systems, with Finland being an exception to this trend. These observations highlight the critical need for developing context-specific funding policies considering local conditions and priorities ([11]). A comprehensive study focusing on universities around the Mediterranean and Black Sea regions utilized advanced analytical procedures to classify various countries into performance clusters. This research demonstrated a notable academic superiority among institutions from Italy, Spain, and France, with the employed clustering metrics providing valuable insights to aid regional development initiatives ([12]). Global university ranking systems foster healthy competition among academic institutions ([13]). A prime example is the Times Higher Education (THE) rankings, meticulously designed to understand universities' teaching and learning environments comprehensively. This study delves into the evolving trends in university impact rankings assessed by THE, explicitly focusing on institutions in East Java, Indonesia. Using a mini-view methodology based on information from THE's official website, this research shows that three key indicators—teaching effectiveness, the overall research environment, and the quality of research output—are consistently prioritized in university impact rankings. This trend has been observed across five universities in East Java, highlighting their performance year after year within THE's ranking framework. In an innovative approach, researchers introduced a novel metric called the Academic Gender Equity Index (Academic_GEI). This index quantitatively evaluates gender-neutral academic environments by examining scholarly output and the composition of faculty members. This methodology has been successfully applied in various contexts, including Japan, the United States, and the European

Union, highlighting existing disparities in gender equity and informing policy initiatives to improve gender balance within academic institutions ([14]). Moreover, a promising approach to improving university rankings involves developing criteria that reduce reliance on conventional indicators ([15]). This study analyzes national and global ranking systems, focusing on the relevance of specific indicators to the educational needs of various countries. The research includes two online focus groups with ten diverse participants to gather insights. These discussions were transcribed and analyzed thematically. In another significant aspect of the research, machine learning techniques like cosine similarity were utilized to investigate the relationship between theoretical frameworks presented in doctoral capstone projects and their subsequent impact on university rankings. The findings reveal that the alignment of theoretical frameworks is crucial in shaping external evaluations of academic institutions. However, the researchers concluded that further interdisciplinary studies are necessary to deepen the understanding of these relationships and their implications for academic performance assessments ([16]).

The comprehensive review highlights the evolving landscape of university ranking methodologies and the critical role of advanced algorithms, regional considerations, and interdisciplinary research in shaping future ranking systems, as shown in Table 1.

Table 2 Comparative analysis of advanced methodologies for university ranking systems.

Authors	Method	Advantages	Disadvantages
Weihua et al., 2025	PROMETHEE with Rough Set Theory	Quantitatively ranks features and resolves multi-source ordered decision-making challenges.	Computationally intensive for large datasets.
Ionut & Nicolae, 2024	Cosine Similarity for Theoretical Frameworks	Highlights discrepancies in theoretical framework usage across ranking groups.	Limited by reliance on U.S. datasets, it lacks global representativeness.
Leslie et al., 2024	Pearson Correlation & Chi-Square Analysis	Identifies key ranking factors such as student-faculty ratio and academic citations.	The results may not be generalized globally if the study is focused on Canadian universities.
Zhenhua et al., 2024	Graph Neural Networks (GNNs)	Constructs preference relation graphs, enhancing ranking accuracy and efficiency.	Requires extensive computational resources.

Haoyu et al., 2024	Ranking Improved Teaching-Learning-Based Optimization (RITLBO)	Identifies critical parameters for diverse applications, demonstrating versatility in ranking-related tasks.	Limited application in educational contexts requires parameter tuning.
Wang et al., 2015	CCRank (Cooperative Coevolutionary Ranking Framework)	Enhances computational efficiency and scalability using evolutionary algorithms.	Requires decomposition of complex ranking challenges into sub-problems.
Pavel, 2015	Comparative Studies (THE, QS, ARWU)	Provides insights into varying methodologies and implications for stakeholders.	Limited alignment between ranking methodologies and local needs.
Fasih et al., 2025	TOPSIS (Multicriteria Optimization Model)	Promotes equitable representation and regional development in academic rankings.	May oversimplify regional complexities and disparities.
Tamada et al., 2023	Cosine Similarity in Doctoral Capstone Analysis	Investigates the relationship between theoretical frameworks and university rankings.	Limited interdisciplinary application; requires broader datasets.

According to Table 2, the research on university ranking systems reveals several limitations. While PROMETHEE combined with rough set theory is effective, it is computationally intensive and requires significant preprocessing for large datasets. Cosine similarity, used for analyzing doctoral capstone projects, is limited by its relevance to U.S. university data, which restricts global applicability and interdisciplinary analysis. Similarly, Pearson correlation and Chi-Square methods may not generalize to global contexts. Graph Neural Networks (GNNs) face scalability issues due to high computational demands, especially on smaller datasets. The Ranking Improved Teaching-Learning-Based Optimization (RITLBO) method is sensitive to initial conditions and parameters. CCRank's approach of breaking down ranking issues into sub-problems struggles with interconnected systems, while comparisons of global ranking systems like THE, QS, and ARWU often overlook local socioeconomic factors. The Academic Gender Equity Index is constrained by data availability, and multicriteria optimization models such as TOPSIS tend to oversimplify regional disparities. Machine learning methods, while improving predictive accuracy, are computationally intensive and reliant on fine-tuning. Overall, these limitations show the need for context-aware methodologies to enhance the reliability and inclusiveness of university ranking systems.

This paper comprehensively analyzed the various advanced factors that influence university rankings, employing a hybrid feature selection approach grounded in machine learning methodologies for model estimation. The primary contributions of this study can be summarized as follows:

1. A Max-Min normalization preprocessing technique was proposed to enhance the accuracy and reliability of the optimal subset selection for university model prediction. This approach scales the features within a specified range, making it easier to compare and analyze different attributes in the dataset (Section 2.1.1).

2. We employed a combination of feature weighting index by ANOVA F-test techniques and artificial neural network for feature selection. Then, we comprehensively compared the performance among three distinct base models: linear regression, random forest, and multilayer perceptron. All models were evaluated using the same training and testing datasets. The objective was to identify which hybrid feature selection techniques yielded the most significant improvements in the predictive performance of the models, ultimately providing insights into the effectiveness of each method (Section 2.1.2).

3. We performed a comparative analysis of the prediction models for university rankings that integrated hybrid feature selection in the model estimation. This evaluation employed a systematic approach using seven performance metrics: accuracy, precision, recall, F-score, mean absolute error (MAE), root mean square error (RMSE), and R-squared (R^2). We analyzed these metrics to determine the hybrid feature selection methods based on machine learning techniques and their impact on model performance (Section 3).

4. We aim to develop a robust model estimation framework that utilizes hybrid feature selection, ensuring practical applicability across various years and diverse data environments. The results of this research are expected to provide a solid foundation for the automated detection and prediction of university ranking trends, ultimately enhancing comfort regulation in this domain and contributing to the development of intelligent prediction systems.

2. Materials and Methods

2.1. Data Sources and Preprocessing

This study presents an innovative approach to predicting university rankings through a hybrid feature selection method within the model estimation. The comprehensive framework, illustrated in Fig. 1, encompasses the critical steps of identifying key variables and constructing, followed by evaluating the estimation model.

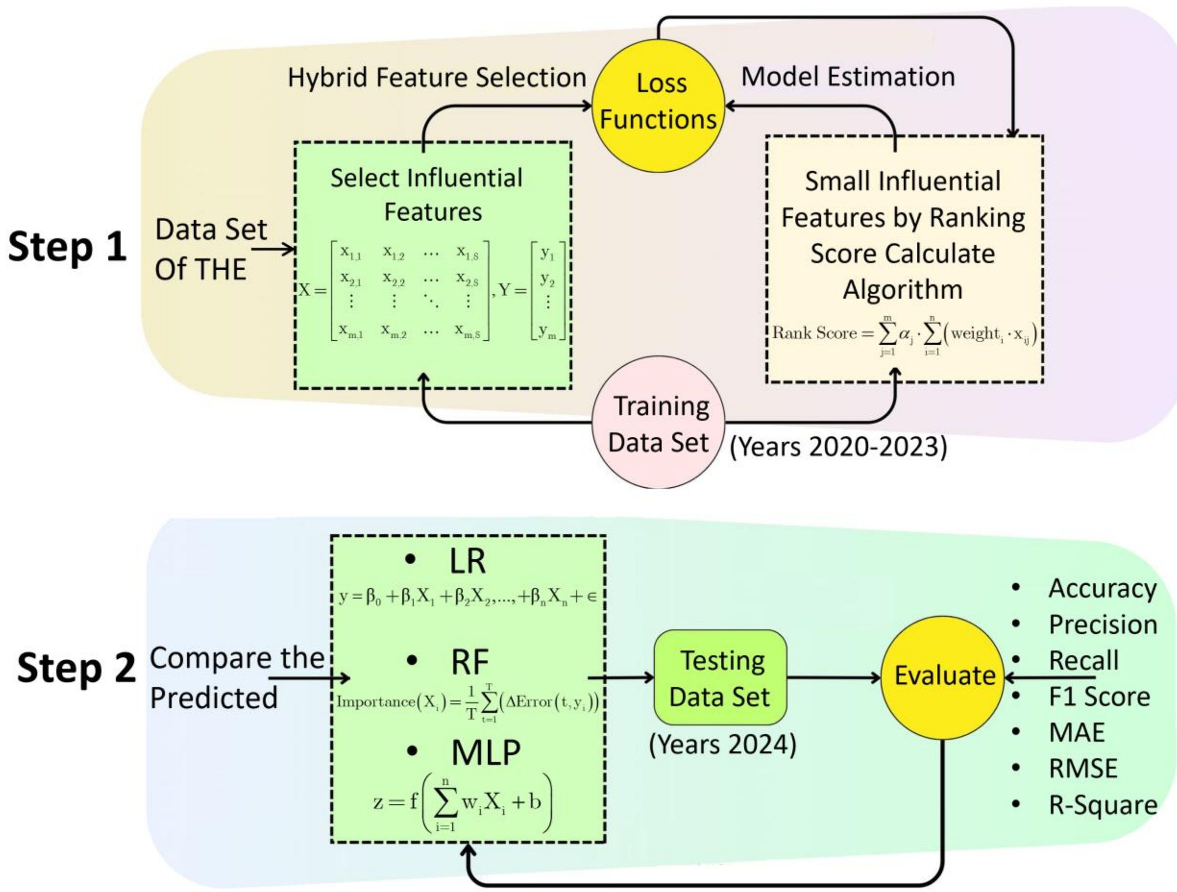

Fig 1. The overall framework of university ranking prediction.

The raw data utilized in this research is sourced from THE University, an online, open-source database known for its extensive collection of university ranking information. This database integrates results from various university rankings conducted between 2021 and 2024, encompassing approximately 1,904 distinct rankings. Most rankings incorporate detailed metrics related to several essential variables contributing to a university's overall assessment. In light of the release of THE database, many research initiatives have concentrated on analyzing specific hyperparameters or models, assessing subjective metrics, and exploring integration with machine learning algorithms ([17]). Nevertheless, the process of variable selection has often been constrained by factors such as the availability of application data, the size of the training dataset, and practical implementation considerations. This study identified eight key variables for inclusion in the analysis. These variables are teaching quality, research performance, citation impact, income generation, international outlook, student population, student-to-staff ratio, and gender ratio (female-to-male) ([18]). Additionally, the study assigns a weight to each variable to reflect its significance in the ranking process. For a comprehensive view of these variables, Table 2 provides an in-depth description sourced from THE university ranking database, elaborating on the definitions and relevance of each variable within the context of university ranking methodologies ([19]).

Table 2 The detailed attributes contained in THE university rankings 2024.

Number	Features	Descriptions
1	Teaching	This indicator reflects the quality of teaching and learning environments at each university, often assessed through metrics such as faculty-student ratios, teaching reputation surveys, and academic infrastructure. A high teaching score signals a commitment to student success and effective pedagogy, which is key to achieving higher ranks.
2	Research	Research output and impact are essential components of university rankings. This score reflects the quantity and quality of research an institution produces and external funding for research activities. Publications, grants, and collaborations typically measure research performance.
3	Citations	Citations serve as a proxy for research impact, indicating how frequently other scholars reference a university's publications. This metric captures the global influence of a university's research, with a higher citation score suggesting that the institution's work contributes significantly to advancing knowledge.
4	Income	This attribute measures financial engagement with industry partners, highlighting a university's ability to attract funding from non-academic sources. A higher industry income score reflects the institution's success in collaborating with industry and generating practical, market-relevant innovations.
5	International Outlook	This score reflects the extent of a university's global engagement, including the diversity of its student and faculty populations, international partnerships, and collaborative research. Universities with a strong international outlook tend to attract a more diverse community and have broader global influence.
6	Number of Students	This attribute reflects the overall size of the student body, providing insights into the scale of operations at the institution. More prominent universities may have more diverse academic offerings, while smaller institutions may focus on specialized areas.
7	Student-to-Staff	This metric captures the teaching intensity at a university by indicating how many students each staff member serves. A lower ratio often correlates with better academic support and personalized learning experiences, contributing to higher teaching scores.
8	Female-to-Male	This demographic attribute reflects gender diversity at the university. Institutions with balanced gender ratios demonstrate inclusivity and diversity, which may positively influence their reputation and ranking.

Data preprocessing is an essential and foundational step in the machine learning pipeline. This process takes raw data from various sources and transforms it into a structured and organized format conducive to developing accurate predictive models. One of the primary activities involved in data preprocessing is data standardization. This consists of normalizing or scaling data points to a consistent range, which can significantly enhance the performance of many algorithms sensitive to input data's magnitude. For example, techniques such as Min-Max scaling and Z-score are commonly employed to ensure that features contribute equally to the model's predictions. In addition to standardization, addressing data imbalance is another critical aspect of preprocessing. Certain classes or categories may be underrepresented in many real-world datasets, leading to biased model predictions. Techniques such as resampling (either oversampling the minority class or undersampling the majority class), using synthetic data generation, or employing algorithmic approaches focusing on misclassified data can help mitigate these issues. Effective data preprocessing also ensures the dataset is compatible with various requirements and constraints of different normalization techniques. By preparing the data thoughtfully, practitioners facilitate meaningful model comparisons, allowing them to evaluate performance across different approaches accurately.

2.1.1 Normalized University Ranking Data

To begin with, we identified and selected eight essential key variables that would contribute significantly to our analysis. We examined the dataset for any apparent outliers or missing values during this process. Rows containing such irregularities were systematically removed to enhance the overall quality of the dataset. This careful cleaning step was crucial in preserving the original dimensions of the input data, ensuring that the model could effectively process and analyze these selected features without distortion. Following the data cleaning, we moved on to the data standardization phase ([20]). Given the essential differences in the distribution and units of each variable, we selected the Max-Min normalization method. This decision ensured that all variables would be on a comparable scale, enabling a more accurate modeling process. This extraction was followed by a thorough cleaning to eliminate any remaining discrepancies. Finally, we applied the normalization process to scale all the values to a standardized range between 0 and 1, as represented in Eq. (1).

$$\mathrm{x}_{\mathrm{norm}} = \frac{\mathrm{x} - \mathrm{x}_{\mathrm{min}}}{\mathrm{x}_{\mathrm{max}} - \mathrm{x}_{\mathrm{min}}} \tag{1}$$

where x is the original value of the feature, x_{norm} is the normalized value of the feature, x_{min} is the minimum value of the feature, and x_{max} is the maximum value of the feature in the dataset. The Max-Min normalization technique proved to be highly effective in addressing the scale disparities between different features in a dataset. By applying this method, we ensured that every variable contributed equally during model training, allowing for a more balanced analysis. Compared to Z-score normalization methods, the Max-Min approach was particularly well-suited for the university ranking database, significantly enhancing the accuracy and stability of our model's predictions. We successfully identified and retained 1,904 qualified samples through the data cleaning and processing efforts. Table 3 details these samples and showcases the results of our normalization efforts and their impact on our analysis. The distribution and the relative values of each feature and the heat map display the relative values of each metric across universities, as shown in Fig. 2.

Table 2 The data using Min-Max normalization is scaled from 0 to 1.

Rank	Teaching	Research	Citations	Income	International Outlook	Number of Students	Student-to-Staff	Female-to-Male
1	0.9732142857142857	1.0	0.9605399792315681	0.9845971563981043	0.984280532043531	0.04502441369819861	0.0448818898	0.0157894737
2	1.0	0.9769392033542976	0.9989615784008307	1.0	0.8573155985489723	0.03003423720664185	0.094488189	0.0263157895
3	0.9955357142857142	0.9601677148846961	1.0	1.0	0.939540507859734	0.02292151970090111	0.0220472441	0.0578947368
4	0.9854910714285715	0.9989517819706499	0.9968847352024922	0.8127962085308057	0.9032648125755743	0.04150121446749953	0.0299212598	0.052631579
5	0.9642857142857143	1.0	0.9823468328141225	0.8566350710900474	0.9830713422007258	0.042568536587387774	0.0496062992	0.0210526316
6	0.9698660714285714	0.9779874213836479	0.9906542056074765	0.9419431279620852	0.8827085852478841	0.0160160492087309	0.0165354331	0.0263157895
7	0.9732142857142857	0.9790356394129979	0.9605399792315681	1.0	0.9008464328899638	0.004590521350587338	0.070866142	0.0789473684
8	0.9095982142857143	0.9528301886792453	0.9885773624091381	0.8921800947867299	0.9939540507859734	0.04196752024803323	0.0519685039	0.0526315789
9	0.8683035714285714	0.9874213836477987	0.9927310488058151	0.9928909952606636	0.8548972188633616	0.0828283414435998	1	0
10	0.9441964285714286	0.9465408805031447	0.9792315680166147	0.840047393364929	0.8016928657799277	0.024660322144757895	0.0448818898	0.0157894737
…	…	…	…	…	…	…	…	…
1904	0.03459821428571428	0.06918238993710692	0.18899273104880585	0.3850710900473933	0.12454655380894798	0.060731664856709416	0	0.052631579

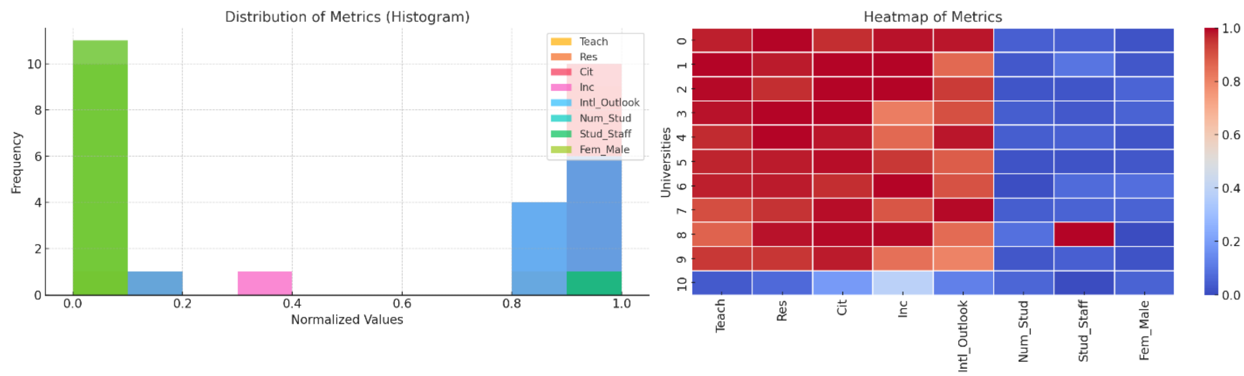

Fig. 2 shows the distribution and relative values of each feature, (a) the histogram shows the distribution of normalized values for all metrics, giving an overview of their spread across universities. (b) the heat map displays the relative values of each metric across universities, with color intensity representing the magnitude of scores.

2.1.2 Feature Selection Strategies

The feature selection process was conducted using a comprehensive two-step approach designed to enhance model performance efficiency and accuracy ([21]).

Step 1 Feature Rankings and Scores: In the initial phase, we employed a combination of feature weighting index and ANOVA F-test techniques for feature selection ([22]). The dataset regarding university rankings comprises eight specific features, each representing essential performance indicators that assess various aspects of a university's quality and impact. These features may include teaching, research, citations, income, international outlook, number of students, student-to-staff, and female-to-male, contributing to a university's overall performance. Our research has analyzed the influence of these performance indicators over the past five years, focusing on their effects across different countries. This country-wise analysis aims to uncover patterns and trends in how these features correlate with university rankings in diverse educational landscapes. To achieve this, we employed ANOVA F-test techniques. This statistical method allows us to determine whether statistically significant differences exist between the means of different groups. The F-score is calculated as Eq. (2) for a single feature.

$$F = \frac{\text{Variance Between Groups (Explained Variance)}}{\text{Variance Within Groups (Unexplained Variance)}} \tag{2}$$

where Explained Variance is a measure of the feature that contributes to separating the groups of the target variable as Eq. (3).

$$\text{Explained Variance} = \frac{\sum_{k=1}^{K} n_k \left(\bar{y}_k - \bar{y}\right)^2}{K - 1} \tag{3}$$

where K is the number of groups (categories of the target variable), n_k is the number of samples in group k, $\bar{y}_k$ is the mean of the target variable for group k, and $\bar{y}$ is the overall target variable. The Unexplained Variance measures the variance of the target variable within each group and is defined as Eq. (4).

$$\text{Unexplained Variance} = \frac{\sum_{k=1}^{K}\sum_{i=1}^{n_k}\left(\bar{y}_{ki} - \bar{y}_k\right)^2}{N - K} \tag{4}$$

where N is the total number of samples. In this phase, we focused on refining our feature selection process by identifying the highest-ranked features derived from the ranked feature list established in Step 1. The feature selection above (Table 3) with a high F-score for the feature significantly contributes to explaining the variance in the target. For example, the computation on the "Teaching" feature is as follows:

1. Calculate group means and overall mean
   - Group 1 (Accept): $\bar{y}_1 = 0.2179$
   - Group 2 (Reject): $\bar{y}_2 = 0.2210$
   - Overall Mean ($\bar{y}$): $\bar{y}_1 = 0.2194$
2. Between-group Variance = 0.0048
3. Within-group = 46.2378
4. Compute F-value by $F = \dfrac{0.0048/(2-1)}{46.2378/(N-2)} = 0.1961$

This computation confirms the F-value for the "Teaching" feature as approximately 0.196, matching the earlier result. The ANOVA F-test results for feature selection are presented in Table 4. Each feature's F-value indicates its significance in explaining the variance in the target variable.

Table 4 ANOVA F-test results for feature selection.

Rank	Feature	F-Value
x_1	Teaching	0.596077198109346
x_2	Research	0.806733351380613
x_3	Citations	1.082245039113657
x_4	Income	0.487751479292153
x_5	International Outlook	0.245189761893036
x_6	Number students	0.128827721924144
x_7	Female-to-Male	0.025521872849439
x_8	Student-to-Staff	0.045977165227463

Step 2 Feature Selected: An Artificial Neural Network (ANN) is utilized to prioritize and rank features based on their significance and contribution to the analysis ([23]). For a dataset with n features, the input layer consists of n neurons. Assume n = 8 features $\{x_1, x_2, x_3, \ldots, x_8\}$, and a target variable y. The dataset is computed as Eq. (5).

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,8} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,8} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m,1} & x_{m,2} & \cdots & x_{m,8} \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \tag{5}$$

where $x_{i,j}$ is the value of the j^{th} feature for the i^{th} sample, and m is the total number of samples. Then, normalize the features X to have zero mean and unit variance using Eq. (6).

$$x_{i,j}^{\text{normalized}} = \frac{x_{i,j} - \mu_j}{\sigma_j} \tag{6}$$

where μ_j is the mean of the j^{th} feature and σ_j is the standard deviation of the j^{th} feature. Next, permutation importance evaluates the impact of a feature by shuffling its values. The logic for computing the model baseline performance is Eq. (7).

$$\text{Baseline Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Samples}} \tag{7}$$

This stage Baseline Accuracy = 0.85. For feature x_j, shuffle its values across all samples, $X_{\text{shuffled}}^{(j)}$ and compute the new accuracy using Eq. (8).

$$\text{Shuffled Accuracy}_j = \frac{\text{Correct Predictions}\left(\text{after shuffling } x_j\right)}{\text{Total Samples}} \tag{8}$$

After shuffling x_1, Shuffled Accuracy_1 = 0.75, and Shuffled Accuracy_2 = 0.83. Next, calculate the importance score for feature x_j using Eq. (9).

$$\text{Importance}_j = \text{Baseline Accuracy - Shuffled Accuracy}_j \tag{9}$$

At this stage, Importance_1 = 0.85 – 0.75 = 0.10 and Importance_2 = 0.85 – 0.83 = 0.02. Higher importance scores indicate that the feature is critical for the model's performance. After computing importance scores for all n features, the feature ranking is based on selecting the top 6 features by Eq. (10).

$$\text{Top Feature} = \left\{x_{j1}, x_{j2}, \ldots, x_{j6}\right\}, \text{ where } \text{Importance}_{j1} > \text{Importance}_{j2} > \ldots \tag{10}$$

At this stage, features are sorted by their important scores: $X_{top6} = [x_3, x_2, x_1, x_4, x_5, x_6]$. To refine our prediction model, we recognized that features such as research, citations, teaching, international outlook, income acceptance rate, and the ratios of female-to-male students and student-to-staff remained constant across different rankings, as illustrated in Table 4.

Table 3 ANN feature importance results.

Rank	Feature	Baseline Accuracy	Shuffled Accuracy	Feature Importance	Selected
1	Research	0.85	0.83	0.9999999999999	**TRUE**
2	Citations	0.85	0.82	0.8999999999999	**TRUE**
3	Teaching	0.85	0.80	0.7999999999999	**TRUE**
4	Income	0.85	0.79	0.6999999999999	**TRUE**
5	International Outlook	0.85	0.78	0.5999999999999	**TRUE**
6	Number students	0.85	0.76	0.5099999999999	**TRUE**
7	Student-to-Staff	0.85	0.74	0.3000000000000	FALSE
8	Female-to-Male	0.85	0.73	0.2000000000001	FALSE

2.1.3 University Ranking Prediction Model

In developing our prediction model, we initiated the process by splitting the ranking dataset into two distinct sets: one for training and the other for testing. To enhance the accuracy of our predictions, we employed a hybrid features selection method to classify the various attributes related to performance. Next, we devised a ranking score calculation for each performance feature. This involved utilizing historical scores from the previous year, which were weighted by their recency, to generate new scores for our predictive analysis. Subsequently, we computed a total score that aggregated the individual performance feature scores, assigning weights based on how significantly each feature influences the overall university ranking.

Consider a university that has systematically collected and analyzed performance scores related to six specific performance features over five consecutive years. These features have been evaluated based on established criteria, providing valuable insights into the university's performance. The weights assigned to each feature for every year indicate their relative importance in the overall assessment. This detailed information is thoroughly outlined in Table 5, which serves as a key reference for understanding trends and changes in performance over time.

Table 5 The information on features and yearly weights of THE dataset.

Features Weights	**Yearly Weights** ($weight_i$)
Research (α_1) = 0.25	Year 2024 = 0.4
Citations (α_2) = 0.20	Year 2023 = 0.3
Teaching (α_3) = 0.20	Year 2022 = 0.2
International Outlook (α_4) = 0.15	Year 2021 = 0.1
Income (α_5) = 0.10	Year 2020 = 0.1
Number of students (α_6) = 0.10	
The weighted sum to 1: 0.25+0.20+0.20+0.15+0.10+0.10=1.00	

Initially, we focused on the university ranking dataset provided by THE. To identify the most significant performance features, we thoroughly analyzed the year-by-year variation in scores across various universities. For example, in building and evaluating our predictive model, we strategically divided the dataset: the data from 2020 to 2023 was designated as the training set, while the data from 2024 was reserved for testing purposes, as shown in Table 6.

Table 6 The details of scores for a university over 5 years.

Year	Citations	Research	Teaching	International Outlook	Income	Number of Students
2024	90	85	80	75	70	60
2023	92	88	82	78	72	62
2022	89	84	81	76	71	61
2021	91	86	83	77	73	63
2020	90	87	85	79	74	65

To train the dataset, we implemented our innovative hybrid feature selection method to detect outliers in each performance feature from 2020 to 2023. Subsequently, we calculated the predicted scores for six specific features, employing the formula outlined in Eq. (11). This systematic approach allowed us to enhance the accuracy of our predictions and gain deeper insights into the factors influencing university rankings ([24]).

$$\text{Rank Score} = \sum_{j=1}^{m} \alpha_j \cdot \sum_{i=1}^{n} \left(\text{weight}_i \cdot x_{ij}\right) \tag{11}$$

where m is the number of performance features, n is the number of years, x_{ij} is the performance score of the j^{th} feature in the i^{th} year, α_j is the weight of the j^{th} feature, and weight_i is the weight for i^{th} year. After that, we generated a total predicted rank score based on the weight of each performance feature as follows:

- $$\begin{aligned}\text{Research Weighted Score} &= 0.25 \cdot \left[(0.4 \cdot 85) + (0.3 \cdot 88) + (0.2 \cdot 84) + (0.1 \cdot 86) + (0.1 \cdot 87)\right] \\ &= 0.25 \cdot \left[34 + 26.4 + 16.8 + 8.6 + 8.7\right] = 0.25 \cdot 94.5 = 23.63\end{aligned}$$
- $$\begin{aligned}\text{Citations Weighted Score} &= 0.20 \cdot \left[(0.4 \cdot 90) + (0.3 \cdot 92) + (0.2 \cdot 89) + (0.1 \cdot 91) + (0.1 \cdot 90)\right] \\ &= 0.25 \cdot \left[36 + 27.6 + 17.8 + 9.1 + 9.0\right] = 0.20 \cdot 99.5 = 19.90\end{aligned}$$
- $$\begin{aligned}\text{Teaching Weighted Score} &= 0.20 \cdot \left[(0.4 \cdot 80) + (0.3 \cdot 82) + (0.2 \cdot 81) + (0.1 \cdot 83) + (0.1 \cdot 85)\right] \\ &= 0.20 \cdot \left[32 + 24.6 + 16.2 + 8.3 + 8.5\right] = 0.20 \cdot 89.6 = 17.92\end{aligned}$$
- $$\begin{aligned}\text{Outlook Weighted Score} &= 0.15 \cdot \left[(0.4 \cdot 75) + (0.3 \cdot 78) + (0.2 \cdot 76) + (0.1 \cdot 77) + (0.1 \cdot 79)\right] \\ &= 0.15 \cdot \left[30 + 23.4 + 15.2 + 7.7 + 7.9\right] = 0.15 \cdot 84.2 = 12.63\end{aligned}$$

- $$\begin{aligned}\text{Income Weighted Score} &= 0.10\cdot\left[(0.4\cdot 70)+(0.3\cdot 72)+(0.2\cdot 71)+(0.1\cdot 73)+(0.1\cdot 74)\right]\\ &= 0.10\cdot[28+21.6+14.2+7.3+7.4] = 0.10\cdot 78.5 = 7.85\end{aligned}$$
- $$\begin{aligned}\text{Student Weighted Score} &= 0.10\cdot\left[(0.4\cdot 60)+(0.3\cdot 62)+(0.2\cdot 61)+(0.1\cdot 63)+(0.1\cdot 65)\right]\\ &= 0.10\cdot[24+18.6+12.2+6.3+6.5] = 0.10\cdot 67.6 = 6.76\end{aligned}$$

Next, calculate the overall rank score by applying the formula outlined in Eq. (12). This will involve substituting the relevant values into the formula to derive a comprehensive score reflecting the earlier ranking criteria. Ensure that all necessary data points are accurately gathered and analyzed to achieve a precise total rank score.

$$\begin{aligned}\text{Rank Score} &= \text{Sum of all weighted scores}\\ &= 23.63+19.90+17.92+12.63+7.85+6.76 = 88.69\end{aligned} \tag{12}$$

Therefore, the predicted university ranking score for the university is 88.69.

2.1.4 Machine Learning Methods

1) Linear regression is a statistical method that models the relationship between a dependent variable, “scores_overall_rank,” and multiple independent variables, like teaching quality, research output, and citation numbers ([25]). It assumes a linear relationship, meaning changes in the independent variables are associated with proportional changes in the dependent variable. A mathematical equation (Eq. (13)) formalizes this relationship, detailing how each independent variable predicts the overall rank.

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2, \ldots, + \beta_n X_n + \in \tag{13}$$

where y is the predicted rank, β_0 is the intercept, β_i is the coefficient (weight) of feature Xi, and ε represents the residual error. The weights (β_i) are computed by minimizing the squared error sum and defined as Eq. (14).

$$\min \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2 \tag{14}$$

The weights (β_i) represent the contribution of each feature to the target rank. In our analysis, these coefficients serve as feature weights.

2) A random forest is a powerful machine-learning method that combines multiple decision trees to enhance predictive performance and robustness ([26]). It evaluates the importance of each feature by calculating the average reduction in prediction error that occurs when a specific feature is utilized across all the decision trees in the ensemble. This feature importance measure provides insights into which variables impact the model’s predictions most, thereby aiding in feature selection and interpretation. The importance score of a feature X_i is defined in Eq. (15).

$$\text{Importance}(X_i) = \frac{1}{T}\sum_{t=1}^{T}\left(\Delta\text{Error}(t, y_i)\right) \tag{15}$$

where T is the number of trees in the forest, ΔError(t, X_i) is the change in prediction error when feature X_i is removed from tree t. The importance score indicates a feature contribution to the model's accuracy across all trees.

3) A Multilayer Perceptron (MLP) is an artificial neural network with multiple layers of interconnected neurons ([27]). The input layer receives data, hidden layers process it, and the output layer produces predictions. During training, the MLP learns to adjust the weights of connections through back-propagation to minimize the error between predictions and target values using a quantitative error expression (Eq. (15)).

$$z = f\left(\sum_{i=1}^{n} w_i X_i + b\right) \tag{16}$$

where z is the neuron output, w_i is the weight for feature X_i, b is the bias term, and f is the active function. During training, the weight is adjusted to minimize the loss by using $\text{Loss} = \frac{1}{2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$. Once the training process is complete, the importance of permutation-based features is assessed by analyzing the impact on the model's performance when each feature is randomly shuffled.

For a comprehensive understanding of the model architecture, Table 7 provides an in-depth overview of the key hyperparameters and functions instrumental in constructing these three different machine learning models, as elaborated in Section.

Table 7 Summarizes the best parameters and key equations used in each model to compute the feature weights of university rankings.

Method	Equation	Description and value
LR	$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2, \ldots, + \beta_n X_n + \in$	Scales feature to normalize data before fitting (True), and whether to calculate the intercept for the model (True).
RF	$\text{Importance}(X_i) = \frac{1}{T}\sum_{t=1}^{T}\left(\Delta\text{Error}(t, y_i)\right)$	Number of trees in the forest (100), Maximum depth of each tree (10), minimum samples required to split an internal node (2), minimum samples required at a leaf node (1), and seed for reproducibility (42).
MLP	$z = f\left(\sum_{i=1}^{n} w_i X_i + b\right)$	Number of neurons in each hidden layer as a tuple (100, 50), activation function for hidden layers (relu), optimization algorithm (adam), regularization term to prevent overfitting (0.0001), maximum number of iterations for convergence (500), and seed for reproducibility (42).

3. Experimental Results

We have structured our data set by allocating the data from 2020 to 2023 for training purposes and reserving the remaining 2024 for testing. We evaluate the effectiveness of our proposed university ranking prediction model. Our analysis identified traditional university ranking prediction techniques, including linear regression, random forest, and multilayer perceptron, to benchmark against our innovative hybrid approach. To thoroughly assess the accuracy and performance of the predictive model, we have employed seven evaluation measures, which will help illustrate how our method stands up to the traditional approaches.

1. Accuracy: The Accuracy score is a standard performance metric used to evaluate the correctness of predictions ([28]). It measures the proportion of correctly predicted outcomes to the total number of observations. In prediction problems, accuracy can also be calculated using the concepts of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). These terms describe the outcomes of predictions compared to actual values as defined in Eq. (17).

$$\text{Accuracy} = \frac{\text{Correct predictions } (\text{TP}+\text{TN})}{\text{Total number of predictions } (\text{TP}+\text{FP}+\text{TN}+\text{FN})} \tag{17}$$

where TP is a correctly predicted positive outcome, TN is a correctly predicted negative outcome, FP is an incorrectly predicted positive outcome, FN is an incorrectly predicted negative outcome, TP+TN represents the total number of correct predictions made by the model, FP+FN represents the total number of incorrect predictions, TP+TN+FP+FN is the total number of predictions made by the model is the sum of all four categories.

2. Precision is a performance metric used to evaluate the accuracy of positive predictions made by a classification model ([29]). It measures the proportion of true positive predictions (correctly identified positives) out of all predicted positive cases. The formula for precision is Eq. (18).

$$\text{Precision} = \frac{\text{True Positives } (\text{TP})}{\text{True Positives } (\text{TP})+\text{False Positives } (\text{FP})} \tag{18}$$

3. Recall is an important metric that evaluates the effectiveness of a model in identifying all relevant positive instances within a dataset. It measures the proportion of actual positive cases the model correctly predicted ([30]). A specific formula for calculating recall is referenced as Eq. (19).

$$\text{Recall} = \frac{\text{True Positives } (\text{TP})}{\text{True Positives } (\text{TP})+\text{False Negatives } (\text{FN})} \tag{19}$$

4. The F1 Score is a metric that combines precision and recall into a single score ([31]). It is beneficial when the importance of precision (minimizing false positives) and recall (minimizing false negatives) must be balanced. The formula for the F1 Score is Eq. (20).

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{20}$$

5. Mean Absolute Error (MAE) is a metric used to evaluate the accuracy of continuous predictions ([32]). It measures the average magnitude of the errors between predicted and actual values without considering their direction. The formula for the MAE is Eq. (21).

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} \left| y_i - \hat{y}_i \right| \tag{21}$$

where n is the total number of data points, y_i is the actual value for the i^{th} observation, $\hat{y}_i$ is the predicted value for the i^{th} observation, $\left| y_i - \hat{y}_i \right|$ is an absolute error for the i^{th} observation.

6. Root Mean Squared Error (RMSE) is a widely used metric to evaluate the performance of regression models ([33]). It measures the average magnitude of prediction errors, penalizing more significant errors more heavily than smaller ones due to the squaring of differences. RMSE is expressed in the same units as the target variable, making it intuitive and interpretable. The formula for the MAE is Eq. (22).

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2} \tag{22}$$

where $\left(y_i - \hat{y}_i \right)^2$ is a squared error for the i^{th} observation.

7. R^2 (R-Square) is a statistical measure that indicates how well the independent variables explain the variance in the dependent variable ([34]). It is commonly used in regression analysis to evaluate the goodness-of-fit of a model. The formula for the R^2 is Eq. (23).

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{23}$$

where $SS_{res} = \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2$ is the residual sum of squares (unexplained variance), measures the prediction error and $SS_{tot} = \sum_{i=1}^{n} \left(y_i - \bar{y}_i \right)^2$ is the total sum of squares (total variance), measures the variance in the actual data.

We thoroughly compared the experimental results in Table 8 alongside the visual data illustrated in Fig. 3. This analysis aimed to predict the university's potential outcomes. By examining the various metrics and trends displayed in the data, we sought insights that could provide a clearer understanding of the university's performance and future trajectory.

Table 8 Comparison of predictions: Linear regression, random forest, multilayer perceptron, and hybrid methods.

Actual Rank	Linear Regression	Random Forest	Multilayer Perceptron	Hybrid Methods
1	2.2545525452145	1.825	1.93252551225252	1.18
2	3.2279961254355	4.025	3.98582585742533	2.07
3	5.012554152554124	6.042	4.0125251425544255	3.98
4	6.225125214255555	5.025	6.0210221154665621	4.02
5	7.152255425212125	6.982	8.0252145582425525	5.60
6	8.251251415125211	7.021	8.7892241252122551	6.14
7	9.874121425212222	8.028	10.0101141458469413	8.01
8	10.12532951741252	9.021	10.9825121422114122	8.45
9	10.98202125222521	9.922	11.0202115252521029	9.09
…	…	…	…	…
1,000	1,048.0020212521522	1,021.252	1,052.225121284125655	1,011.25

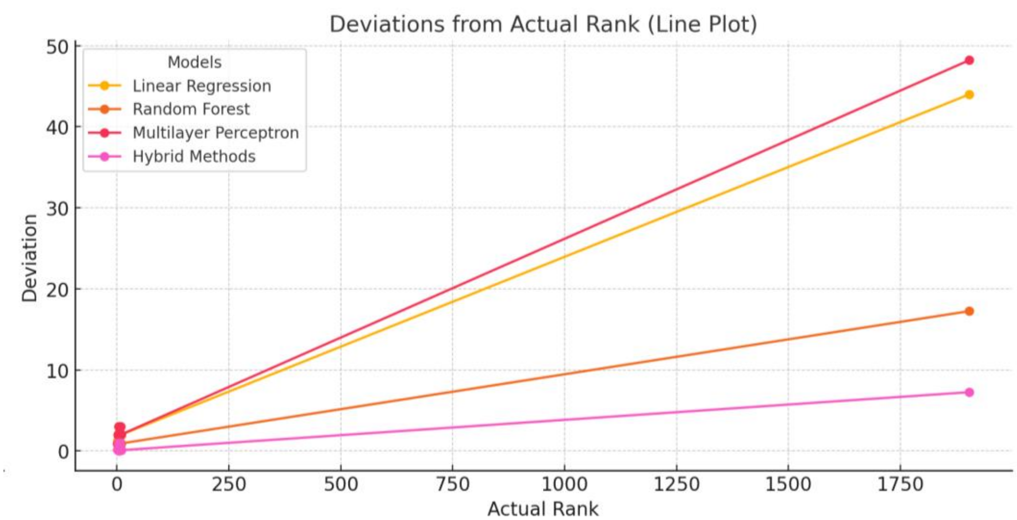

Fig. 3 The graphs provide deeper insights into model performance and deviations.

In Table 8 and Fig. 3, it is evident that hybrid-based methods consistently align more closely with the actual rankings across most of the scenarios examined, highlighting their overall effectiveness in capturing the underlying patterns of the data. In contrast, the predictions made by the random forest and MLP models successfully encapsulate the general trends within the data; however, there are instances where these predictions can diverge from the actual values. This variation points to certain limitations inherent in their predictive capabilities. Furthermore, the linear regression analysis outcomes reveal more significant prediction variability than the hybrid and ensemble methods. This variability underscores the complexities associated with the convergence of the model and the challenges involved in tuning its parameters effectively. Such issues may lead to fluctuations in predictive accuracy, indicating that while linear regression can be helpful, it may struggle to provide consistent results in this context. The performance of various methodologies highlights the complexities of predictive modeling and underscores the need to choose the appropriate approach based on the dataset's specific characteristics.

To compute the performance of our forecasting models using metrics (accuracy, precision, recall, F1 score, MAE, RMSE, and R^2), we need to frame the problem as a classification problem. Forecasting rankings naturally results in continuous numeric predictions, but we can transform the problem into rank categories or binary classes (e.g., Top 1,000 vs. Not Top 1,000) to use classification metrics.

Step 1: Transform the forecast into binary classification: Predict the Top 1,000 universities (Class = 1) vs. Not the Top 1,000 (Class = 0).

Step 2 Compute predictions and compare them to the actual 2024 rankings.

Step 3 Calculate Metrics: Accuracy, Precision, Recall, F1 score, MAE, RMSE, and R^2.

The error indicates that the rank column contains strings instead of numeric values. As displayed in Table 9 and Fig. 4, ensure the data is correctly cleaned by converting the ranks to numeric, handling potential errors, and rerunning the analysis.

Table 9 The performance metrics across Linear Regression, Random Forest, Multilayer Perceptron, and Hybrid Methods.

Metric	Linear regression	Random Forest	Multilayer Perceptron	**Hybrid Methods**
Accuracy	0.8423661	0.911	0.8933521	**0.971**
Precision	0.8454558	0.934	0.8711255	**0.985**
Recall	0.8545588	0.944	0.8714258	**0.971**
F-score	0.8411777	0.959	0.8878852	**0.972**
MAE	1.1825559	0.194	1.1112528	**0.034**
RMSE	1.16255741	0.174	1.14415255	**0.028**
R^2	1.17369699	0.215	1.13454854	**0.053**

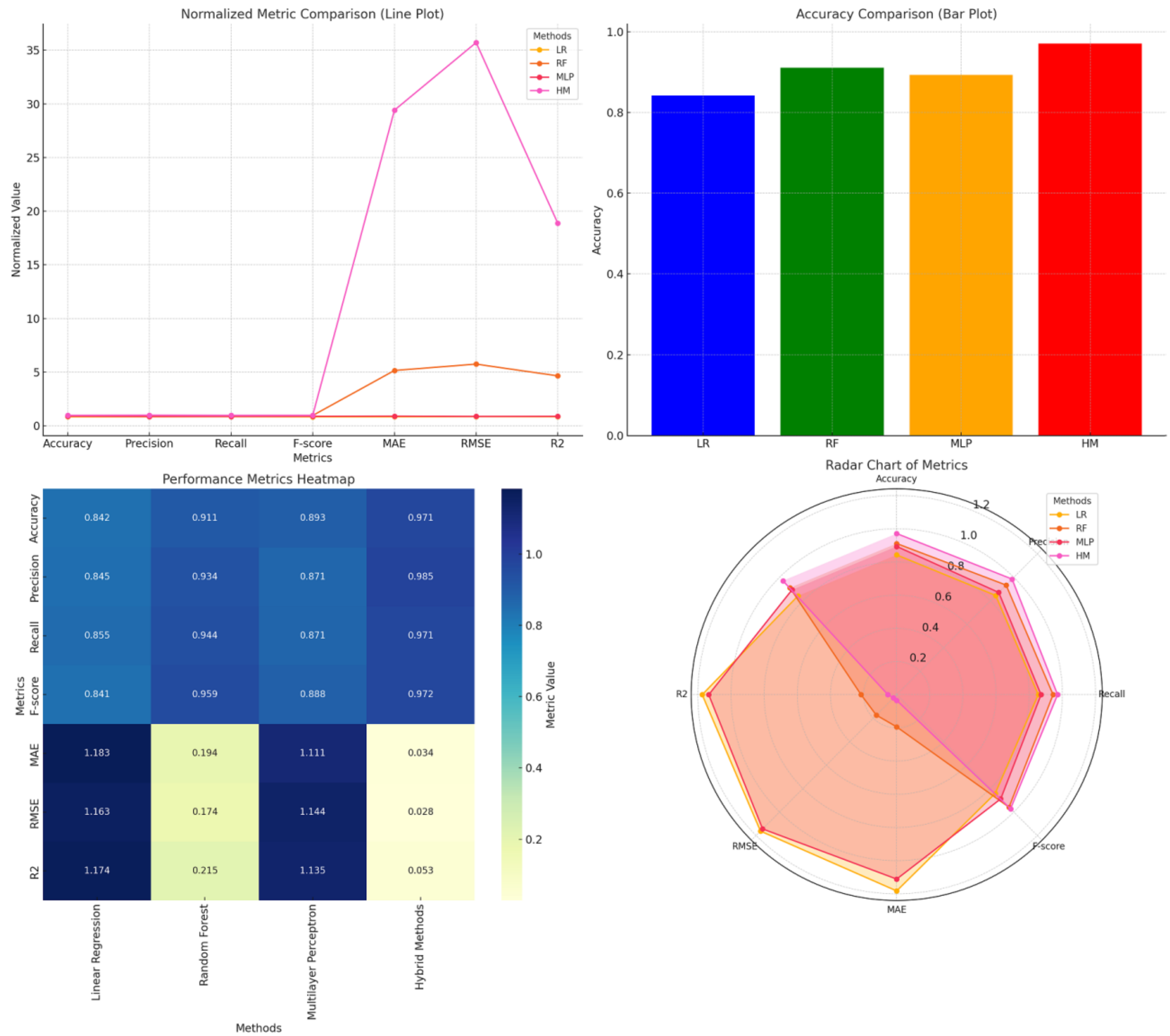

Fig. 4 Comparison of predictive model performance across multiple metrics with Line Plot, Bar Plot, tabular, and radial format.

Fig. 4 provides a detailed comparison of four predictive modeling approaches—linear Regression (LR), Random Forest (RF), MLP, and Hybrid Methods (HM)—across multiple performance metrics, including Accuracy, Precision, Recall, F1-score, MAE, and RMSE. The analysis highlights the strengths and weaknesses of each method, providing a clear understanding of their suitability for predictive tasks. HM consistently demonstrated the best performance across all metrics, achieving the highest accuracy (0.971), precision (0.985), recall (0.971), and F1-score (0.972). These results indicate that HM effectively balances TP and FP predictions while minimizing errors. Additionally, the error metrics for HM were the lowest among all models, with MAE at 0.034 and RMSE at 0.028, further solidifying its superiority in providing highly reliable predictions. The radar chart for HM reflects a nearly perfect, symmetrical shape, indicating consistent performance across all metrics. This suggests that integrating multiple modeling approaches in HM effectively combines their strengths, resulting in a robust solution for complex predictive tasks. RF occurred as the second-best performer, achieving high accuracy (0.911), precision (0.934), recall (0.944), and F1-score (0.959). Though higher than HM (MAE: 0.194, RMSE: 0.174), its error metrics were significantly better than those of LG and MLP. RF's ability to handle non-linear relationships and provide strong predictions makes it a reliable standalone model. The heatmap and radar chart for RF show consistently high values across most metrics, though with slight deviations in precision and error metrics compared to HM. RF represents an excellent choice when computational simplicity is preferred or when combining models in a hybrid approach is not feasible. MLP displayed moderate performance, with an accuracy of 0.893, precision of 0.871, recall of 0.871, and F-score of 0.888. While it outperformed LR, its performance lagged behind RF and HM, particularly in error metrics (MAE: 1.111, RMSE: 1.144). The radar chart for MLP highlights its imbalanced performance across metrics, indicating areas of strength in recall and precision but weaknesses in error reduction. MLP's reliance on neural network architectures provides flexibility and adaptability, but its higher error rates suggest further optimization to achieve performance levels comparable to RF and HM. LR exhibited the weakest performance across all metrics, reaching the lowest accuracy (0.842), precision (0.845), recall (0.854), and F1-score (0.841). Its error metrics were the highest among all models, with MAE at 1.182 and RMSE at 1.163. The radar chart for LR reveals a small and irregular shape, reflecting its significant limitations in handling complex and non-linear relationships. This makes LR unsuitable for tasks requiring high precision or error minimization. While LR is computationally efficient and interpretable, its inability to capture non-linear interactions between variables restricts its utility in advanced predictive modeling.

This study demonstrates the clear advantage of HM for tasks requiring high accuracy, precision, and minimal error. RF serves as a strong alternative, balancing performance and simplicity. MLP, while less competitive, offers the potential for further optimization. In contrast, LR's limitations make it less suitable for this task. These findings highlight the value of leveraging advanced methods like HM to address the complexities of modern predictive tasks, especially where reliability and precision are critical. Future research could explore HM's scalability, computational efficiency, and potential enhancements to RF and MLP for specific applications.

4. Conclusions

This study evaluated the performance of four predictive modeling techniques—LG, RF, MLP, and HM—across multiple evaluation metrics, including Accuracy, Precision, Recall, F-score, MAE, RMSE, and R^2. The results indicate that Hybrid Methods consistently outperformed other approaches, demonstrating their ability to balance predictive accuracy and error minimization. RF also showed strong results as a standalone model, balancing high precision and recall with relatively low error rates. MLP exhibited moderate performance, requiring further optimization to match the efficiency and accuracy of RF and HM. While computationally efficient, LR showed the weakest performance due to its inability to capture complex, non-linear relationships effectively. These findings highlight the superiority of HM for tasks requiring high precision and minimal error and suggest that RF remains a strong alternative for more straightforward implementations. This research emphasizes the importance of leveraging advanced methods to address the complexities of modern predictive tasks.

Future work should focus on enhancing HM by integrating advanced models and optimizing combination strategies to improve robustness and scalability further. Random Forest can benefit from hyperparameter tuning, feature engineering, and dimensionality reduction, while MLP could be improved through architecture optimization, regularization techniques, and transfer learning. Expanding the scope of LR with non-linear extensions or kernel methods could address its limitations. Incorporating explainability techniques like SHAP or LIME would improve interpretability, particularly for complex models. Testing these methods on diverse datasets and evaluating their performance in real-time applications would ensure their adaptability and scalability. Lastly, detailed error analysis and exploring alternative metrics can refine model performance, making these approaches more effective in solving complex, real-world predictive tasks.

Acknowledgments: This research project was financially supported by Mahasarakham Business School, Mahasarakham University, Thailand. The authors express their sincere gratitude to the anonymous reviewers and editors for their insightful critiques and constructive suggestions, which have significantly enhanced the quality and depth of this manuscript. Their invaluable feedback has been instrumental in refining our research methodology, clarifying our findings, and ultimately strengthening the overall impact of this study.

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] E. Tavoletti, F. Şahin, V. Taras, et al. A Configurational Analysis of University Ranking, Cognitive Ability and Cultural Intelligence in Global Virtual Teams: The Role of Academic Pedigree, Int. J. Manag. Educ. 22 (2024), 101040. https://doi.org/10.1016/j.ijme.2024.101040.

[2] W. Xu, Z. Yang, Preference Ranking Organization Method for Enrichment Evaluation-Based Feature Selection for Multiple Source Ordered Information Systems, Eng. Appl. Artif. Intell. 142 (2025), 109935. https://doi.org/10.1016/j.engappai.2024.109935.

[3] I.D. Stanciu, N. Nistor, Doctoral Capstone Theories as Indicators of University Rankings: Insights from a Machine Learning Approach, Comput. Human Behav. 164 (2025), 108504. https://doi.org/10.1016/j.chb.2024.108504.

[4] L.J. Wardley, E. Rajabi, S.H. Amin, M. Ramesh, A Machine Learning Approach Feature to Forecast the Future Performance of the Universities in Canada, Mach. Learn. Appl. 16 (2024), 100548. https://doi.org/10.1016/j.mlwa.2024.100548.

[5] Z. Meng, R. Lin, B. Wu, Graph Neural Networks-Based Preference Learning Method for Object Ranking, Int. J. Approx. Reason. 167 (2024), 109131. https://doi.org/10.1016/j.ijar.2024.109131.

[6] H. Wang, X. Yu, A Ranking Improved Teaching-Learning-Based Optimization Algorithm for Parameters Identification of Photovoltaic Models, Appl. Soft Comput. 167 (2024), 112371. https://doi.org/10.1016/j.asoc.2024.112371.

[7] S. Wang, Y. Wu, B.J. Gao, K. Wang, H.W. Lauw, J. Ma, A Cooperative Coevolution Framework for Parallel Learning to Rank, IEEE Trans. Knowl. Data Eng. 27 (2015), 3152–3165. https://doi.org/10.1109/TKDE.2015.2453952.

[8] V.R. Yeravdekar, G. Tiwari, Global Rankings of Higher Education Institutions and India’s Effective Non-Presence: Why Have World-Class Universities Eluded the Indian Higher Education System? And, How Worthwhile Is the Indian Government’s Captivation to Launch World Class Universities?, Procedia – Soc. Behav. Sci. 157 (2014), 63–83. https://doi.org/10.1016/j.sbspro.2014.11.010.

[9] A.P. Pavel, Global University Rankings - A Comparative Analysis, Procedia Econ. Finance 26 (2015), 54–63. https://doi.org/10.1016/S2212-5671(15)00838-2.

[10] E. Tavoletti, F. Şahin, V. Taras, W. Farrell, L. Florea, A Configurational Analysis of University Ranking, Cognitive Ability and Cultural Intelligence in Global Virtual Teams: The Role of Academic Pedigree, Int. J. Manag. Educ. 22 (2024), 101040. https://doi.org/10.1016/j.ijme.2024.101040.

[11] E. Adam, ‘Governments Base Performance-Based Funding on Global Rankings Indicators’: A Global Trend in Higher Education Finance or a Global Rankings Literature Fiction? A Comparative Analysis of Four Performance-Based Funding Programs, Int. J. Educ. Dev. 76 (2020), 102197. https://doi.org/10.1016/j.ijedudev.2020.102197.

[12] V.M. Moskovkin, E.V. Pupynina, N.P. Zaitseva, R.V. Lesovik, Methodology for Comparative Analysis of University Rankings, with the Mediterranean and Black Sea Region Countries Taken as an Example, Middle-East J. Sci. Res.18 (2013), 1656-1665.

[13] F.B. Ilhami, M. Khasanah, N. Suprapto, B.S. Widodo, Exploring Perspective of Impact Ranking University Trend in Five Universities in East Java Indonesia, Soc. Sci. Human. Open 11 (2025), 101284. https://doi.org/10.1016/j.ssaho.2025.101284.

[14] P.A. Pinto-Delacadena, V. Liern, P. Acosta-Vargas, A. Vinueza-Cabezas, A Multicriteria Approach to Ranking Latin-American Universities Based on Region-Specific Criteria, Technol. Forecast. Soc. Change 208 (2024), 123725. https://doi.org/10.1016/j.techfore.2024.123725.

[15] T. Rafique, M.U. Awan, M. Shafiq, K. Mahmood, Exploring the Role of Ranking Systems towards University Performance Improvement: A Focus Group-Based Study, Heliyon 9 (2023), e20904. https://doi.org/10.1016/j.heliyon.2023.e20904.

[16] K. Tamada, E. Jotaki, N. Tsukamoto, S. Sagara, J.N. Kondo, M. Mori, M. Waga, S. Brown, Evaluation of the Gender-Neutral Academic Climate on Campus for Women Faculty in STEM Fields, Int. J. Educ. Res. Open 7 (2024), 100390. https://doi.org/10.1016/j.ijedro.2024.100390.

[17] J. Li, H. Wang, H. Luo, X. Jiang, E. Li, A Ranking Prediction Strategy Assisted Automatic Model Selection Method, Adv. Eng. Inform. 57 (2023), 102068. https://doi.org/10.1016/j.aei.2023.102068.

[18] I.D. Stanciu, N. Nistor, Doctoral Capstone Theories as Indicators of University Rankings: Insights from a Machine Learning Approach, Comput. Human Behav. 164 (2025), 108504. https://doi.org/10.1016/j.chb.2024.108504.

[19] C. Perez-Esparrells, E. Orduna-Malea, Do the Technical Universities Exhibit Distinct Behaviour in Global University Rankings? A Times Higher Education (THE) Case Study, J. Eng. Technol. Manag. 48 (2018), 97–108. https://doi.org/10.1016/j.jengtecman.2018.04.007.

[20] Y.S. Kim, M.K. Kim, N. Fu, J. Liu, J. Wang, J. Srebric, Investigating the Impact of Data Normalization Methods on Predicting Electricity Consumption in a Building Using Different Artificial Neural Network Models, Sustain. Cities Soc. 118 (2025), 105570. https://doi.org/10.1016/j.scs.2024.105570.

[21] A.M. Dalloo, A.J. Humaidi, Optimizing Machine Learning Models with Data-Level Approximate Computing: The Role of Diverse Sampling, Precision Scaling, Quantization and Feature Selection Strategies, Results Eng. 24 (2024), 103451. https://doi.org/10.1016/j.rineng.2024.103451.

[22] O. Polushkina-Merchanskaya, M.D. Sorochan Armstrong, C. Gómez-Llorente, et al. Considerations for Missing Data, Outliers and Transformations in Permutation Testing for ANOVA with Multivariate Responses, Chemometrics Intell. Lab. Syst. 258 (2025), 105320. https://doi.org/10.1016/j.chemolab.2025.105320.

[23] H. Parastar, P. Weller, Feature Selection and Extraction Strategies for Non-Targeted Analysis Using GC-MS and GC-IMS: A Tutorial, Anal. Chim. Acta 1343 (2025), 343635. https://doi.org/10.1016/j.aca.2025.343635.

[24] M. Taillardat, A.-L. Fougères, P. Naveau, R. De Fondeville, Evaluating Probabilistic Forecasts of Extremes Using Continuous Ranked Probability Score Distributions, Int. J. Forecast. 39 (2023), 1448–1459. https://doi.org/10.1016/j.ijforecast.2022.07.003.

[25] N. Gupta, S. Sivananthan, B.K. Sriperumbudur, Optimal Rates for Functional Linear Regression with General Regularization, Appl. Comput. Harm. Anal. 76 (2025), 101745. https://doi.org/10.1016/j.acha.2024.101745.

[26] L. Zhu, T. Yang, S. Li, F. Yang, C. Jiang, L. Xie, Experimental Investigation and Prediction of Chemical Etching Kinetics on Mask Glass Using Random Forest Machine Learning, Chem. Eng. Res. Design 213 (2025), 309–318. https://doi.org/10.1016/j.cherd.2024.12.014.

[27] I. Shomope, M. Tawalbeh, A. Al-Othman, F. Almomani, Predicting Biohydrogen Production from Dark Fermentation of Organic Waste Biomass Using Multilayer Perceptron Artificial Neural Network (MLP–ANN), Comput. Chem. Eng. 192 (2025), 108900. https://doi.org/10.1016/j.compchemeng.2024.108900.

[28] S.S. Rana, S. Vichare, A. Gupta, A. Rajagopalan, O.P. Kharbanda, Evaluation of the Accuracy of a Simplified Hybrid Method versus the Traditional Digital Method for Measuring Palatal Volume, J. Oral Biol. Craniofac. Res. 15 (2025), 5–10. https://doi.org/10.1016/j.jobcr.2024.10.011.

[29] J. Zhao, Z. Hou, Q. Wang, et al. YOLOrot2.0: A Novel Algorithm for High-Precision Rice Seed Size Measurement with Real-Time Processing, Smart Agric. Technol. 9 (2024), 100599. https://doi.org/10.1016/j.atech.2024.100599.

[30] S.J.J. Leemans, A. Polyvyanyy, Stochastic-Aware Precision and Recall Measures for Conformance Checking in Process Mining, Inf. Syst. 115 (2023), 102197. https://doi.org/10.1016/j.is.2023.102197.

[31] Z. DeVries, E. Locke, M. Hoda, et al. Using a National Surgical Database to Predict Complications Following Posterior Lumbar Surgery and Comparing the Area under the Curve and F1-Score for the Assessment of Prognostic Capability, Spine J. 21 (2021), 1135–1142. https://doi.org/10.1016/j.spinee.2021.02.007.

[32] Y. Tang, L. Shang, R. Zhang, J. Li, H. Fu, Hybrid Divergence Based on Mean Absolute Scaled Error for Incipient Fault Detection, Eng. Appl. Artif. Intell. 129 (2024), 107662. https://doi.org/10.1016/j.engappai.2023.107662.

[33] D.S.K. Karunasingha, Root Mean Square Error or Mean Absolute Error? Use Their Ratio as Well, Inf. Sci. 585 (2022), 609–629. https://doi.org/10.1016/j.ins.2021.11.036.

[34] K. Taira, D. McInnes, L. Zhang, How Many Data Points and How Large an R-Squared Value Is Essential for Arrhenius Plots?, J. Catal. 419 (2023), 26–36. https://doi.org/10.1016/j.jcat.2023.01.033.